

**UNITED STATES DISTRICT COURT  
SOUTHERN DISTRICT OF NEW YORK**

AUTHORS GUILD, et al.,

Plaintiffs,

v.

OPEN AI INC., et al.,

Defendants.

THE NEW YORK TIMES COMPANY,

Plaintiff,

v.

MICROSOFT CORPORATION, et al.,

Defendants.

**Consolidated:**

Case No. 1:23-cv-08292-SHS-OTW

Case No. 1:23-cv-10211-SHS-OTW

**Consolidated:**

Case No. 1:23-cv-11195-SHS-OTW

Case No. 1:24-cv-03285-SHS-OTW

Case No. 1:24-cv-04872-SHS-OTW

**MEMORANDUM OF POINTS AND AUTHORITIES IN SUPPORT OF  
MICROSOFT'S RULE 72(A) OBJECTIONS TO DISCOVERY ORDERS  
REFUSING INQUIRY INTO EVIDENCE OF ECONOMIC EFFECTS AND  
TECHNOLOGY USES RELEVANT TO DEFENSES OF FAIR USE AND  
SUBSTANTIAL NONINFRINGEMENT USES (NYT ECF 351, 354, 355 & AG ECF 289)**

## **TABLE OF CONTENTS**

TABLE OF AUTHORITIES .....	ii
INTRODUCTION .....	1
FACTUAL AND PROCEDURAL BACKGROUND.....	2
A. Large Language Models Are A Profound Advance In Artificial Intelligence and A Powerful Tool For Human Flourishing.....	3
B. Plaintiffs' Claims For Direct and Secondary Copyright Infringement Against Microsoft.....	5
C. Microsoft's Discovery Requests At Issue In The Underlying Motions.....	6
D. The Orders At Issue. ....	9
ARGUMENT .....	10
I. THE ORDER ERRONEOUSLY REFUSES DISCOVERY INTO ECONOMIC EFFECTS RELEVANT TO A KEY FOURTH FACTOR ISSUE: SUBSTITUTION.....	11
II. THE ORDER IMPROPERLY RESTRICTS THE PRODUCTION OF EVIDENCE RELATED TO THE DEVELOPMENT AND USE OF GENERATIVE AI BY PLAINTIFFS. ....	17
A. The Order Fails To Address The Application Of <i>Google v. Oracle</i> To The Relevance Of Evidence Of The Times's Training Of Its Own Model.....	17
B. The Order Improperly Narrows The Scope Of Relevant Uses For The First Factor.....	20
C. The Order Fails To Address Microsoft's Substantial Noninfringing Uses Defense (NYT ECF 354/355 & AG ECF 289).....	24
CONCLUSION.....	25

**TABLE OF AUTHORITIES**

	<b>Page(s)</b>
<b>Cases</b>	
<i>Am. Geophysical Union v. Texaco Inc.</i> , 60 F.3d 913 (2d Cir. 1994).....	12-13, 22
<i>Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith</i> , 598 U.S. 508 (2023).....	21, 22
<i>Authors Guild, Inc. v. HathiTrust</i> , 755 F.3d 87 (2d Cir. 2014).....	12, 16, 20
<i>Authors Guild v. Google, Inc.</i> , 804 F.3d 202 (2d Cir. 2015).....	12, 20
<i>Bill Graham Archives v. Dorling Kindersley Ltd.</i> , 448 F.3d 605 (2d Cir. 2006).....	13
<i>Borjas v. N.Y.C. Dept. of Educ.</i> , No. 23-cv-10829, 2024 U.S. Dist. LEXIS 219579 (S.D.N.Y. Dec. 3, 2024) .....	10
<i>Campbell v. Acuff-Rose Music, Inc.</i> , 510 U.S. 569 (1994).....	11, 12, 15, 16
<i>Cartoon Network LP, LLLP v. CSC Holdings, Inc.</i> , 536 F.3d 121 (2d Cir. 2008).....	24
<i>Concord Music Group, Inc. et al. v. Anthropic PBC</i> , Case No. 5:24-cv-03811 (N.D. Cal.) .....	23
<i>Easley v. Cromartie</i> , 532 U.S. 234 (2001).....	10
<i>EMI Christian Music Grp., Inc. v. MP3tunes, LLC</i> , 844 F.3d 79 (2d. Cir. 2016).....	24
<i>Feist Publications, Inc. v. Rural Telephone Service Co.</i> , 499 U.S. 340 (1991).....	23
<i>Fioranelli v. CBS Broad., Inc.</i> , No. 15-cv-952 (VSB), 2019 WL 1059993 (S.D.N.Y. Mar. 6, 2019) .....	16
<i>Fox News Network, LLC v. TVEyes, Inc.</i> , 43 F. Supp. 3d 379 (S.D.N.Y. 2014).....	20-21

<i>Google LLC v. Oracle Am., Inc.</i> , 593 U.S. 1 (2021).....	<i>passim</i>
<i>Granite State Ins. Co. v. Clearwater Ins. Co.</i> , No. 09 CIV. 10607 RKE, 2012 WL 1520851 (S.D.N.Y. Apr. 30, 2012).....	16
<i>Harper &amp; Row, Publ'rs. v. Nation Enters.</i> , 471 U.S. 539 (1985).....	12
<i>Hedgeye Risk Mgmt., LLC v. Dale</i> , No. 21-cv-3687 ALC RWL, 2023 WL 4235768 (S.D.N.Y. June 28, 2023) .....	10
<i>NXIVM Corp. v. Ross Inst.</i> , 364 F.3d 471 (2d Cir. 2004).....	12
<i>Perfect 10, Inc. v. Amazon.com, Inc.</i> , 508 F.3d 1146 (9th Cir. 2007) .....	20
<i>Sony Corp. of Am. v. Universal City Studios, Inc.</i> , 464 U.S. 417 (1984).....	24, 25
<i>UMG Recordings, Inc. et al. v. Suno, Inc. et al.</i> , Case No. 1:24-cv-11611 (D. Mass.) .....	23
<i>United States CFTC v. Parnon Energy, Inc.</i> , 593 Fed. Appx. 32 (2d Cir. 2014).....	15-16
<i>Weiss v. LaSuisse</i> , 161 F. Supp. 2d 305, 321 (S.D.N.Y. 2001).....	10
<b>Statutes</b>	
17 U.S.C. § 107.....	6, 22
<b>Federal Rules of Civil Procedure</b>	
Fed. R. Civ. P. 26.....	15-16
Fed. R. Civ. P. 72(a) .....	2, 10, 17, 25
<b>Constitutional Provisions</b>	
U.S. Const., Art. I, §8, cl. 8.....	23

## **INTRODUCTION**

These cases before the Court will determine whether groundbreaking new AI technology developed by Microsoft and OpenAI can flourish. There is no dispute that the only path to answering this question is through a thorough analysis of fair use as a defense to Plaintiffs' claims of copyright infringement. Microsoft requests review of the Order (NYT ECF 344) which, in the guise of refusing certain discovery and as applied to Microsoft's separate letter motions, has the effect of materially curtailing Microsoft's ability to fully present the merits of the first and fourth factors of fair use and the separate defense of substantial noninfringing uses.

Magistrate Judge Wang issued ECF 344 to resolve a discovery dispute between OpenAI and The New York Times Company ("The Times"). She incorporated that Order by reference into minute orders denying discovery requested by Microsoft in its subsequent motions to compel. *See* NYT ECF 351, 354 & 355 and AG ECF 289 (denying Microsoft's letter motions NYT ECF 315, 320, 321 and AG ECF 263 "for the reasons stated in [NYT] ECF 344").<sup>1</sup> The distinct evidence sought by Microsoft regarding the economic effects of the alleged infringing technology and the development and uses of that technology is core to the key defenses of fair use and substantial noninfringing uses. In most instances, the discovery at issue sought by Microsoft is of the type that could only be obtained from the Plaintiffs themselves. As a result, by refusing discovery into this evidence, Judge Wang's orders have effectively precluded this evidence altogether, with far-reaching implications for consideration of the merits of the case.

Regrettably, the Order is contrary to law in multiple respects when applied to Microsoft's requests. Mainly, the Order does not address controlling points of law advanced by Microsoft

---

<sup>1</sup> The "News" cases have been consolidated and include *The New York Times Company v. Microsoft Corporation, et al.*, 1:23-cv-11195 (hereafter when referring to the docket, "NYT"), *Daily News LP, et al. v. Microsoft Corporation, et al.*, 1:24-cv-03285, and *The Center for Investigative Reporting, Inc. v. OpenAI, Inc., et al.*, 1:24-cv-04872. The "Class" cases have been consolidated and include *Authors Guild, et al. v. OpenAI Inc., et al.*, 1:23-cv-08292 (hereafter when referring to the docket, "AG") and *Alter, et al. v. OpenAI Inc., et al.*, 1:23-cv-10211.

because it was a decision written analyzing different discovery and separate legal arguments brought by OpenAI. Thus, the Order does not address the importance of the economic evidence sought by Microsoft concerning substitution in Plaintiffs' traditional markets, instead improperly cabining the scope of the fair use factor four inquiry to a licensing market that may or may not even be relevant. Part I, *infra*. The Order also fails to address the lengthy passage in the Supreme Court's decision in *Google v. Oracle* considering the importance to the fourth factor of a plaintiff's efforts to develop the technology at issue. Part II.A, *infra*. Further, the Order does not account for the importance of the pro-copyright benefits of the accused technology, central to consideration of the first fair use factor as well as the wholly separate defense of substantial noninfringing uses, thus rejecting Microsoft's requested evidence regarding Plaintiffs' uses. Parts II.B,C, *infra*.

Microsoft respectfully requests under Rule 72(a) that the Court set aside NYT ECF 351, 354, 355 and AG ECF 289 as both contrary to law and because a definite mistake has been committed in curtailing the merits of the defenses in these cases. The Court should order that all of Microsoft's requested discovery be produced.

#### **FACTUAL AND PROCEDURAL BACKGROUND**

The Plaintiffs in the News and Class cases have brought claims against Microsoft and OpenAI centrally focused on the alleged use of their copyrighted works to train the Large Language Models ("LLMs") that power consumer facing products like ChatGPT and Copilot (formerly known as Bing Chat). Set forth below is the factual background and key aspects of the technology at issue necessary to understand the importance of the requested discovery in establishing Microsoft's affirmative defenses.

**A. Large Language Models Are A Profound Advance In Artificial Intelligence and A Powerful Tool For Human Flourishing.**

Over the past several years, researchers and engineers have made dramatic advances in the field of artificial intelligence. Perhaps the most astounding is the “large language model,” or LLM. As relevant here, an LLM is a machine learning model that can process and produce natural language text. *See* NYT ECF 170 (“NYT Complaint”) ¶¶ 61, 64, 72, 75. Plaintiffs in the Class and News cases have brought claims against OpenAI and Microsoft challenging the use of copyrighted works to train these LLMs, and, in the case of the News Plaintiffs, on the use of those models in next generation search engines that respond to queries with natural language answers. *See generally* AG ECF 69 (“Class Complaint”); NYT Complaint.

Most commercially available LLMs today are trained on vast amounts of data obtained from the internet, the body of which for any particular model is referred to as its “training dataset” or “training corpus.” Because computers cannot inherently understand natural language, the training data must first be represented mathematically. This is accomplished by breaking the text down into small units which are used to represent all elements of language. These “tokens” generally represent less than an entire word.

The revolutionary “transformer” architecture, invented in 2017, along with vast quantities of language data available on the internet, has spurred this recent period of model development. The transformer enables developers to teach the model to pay attention to the varying contexts of the tokens. The model thus learns concepts and principles of human language by acquiring the statistical relationships of tokens to one another across a massive volume, the resulting statistical relationships of the tokens are embedded into the “parameters” of the model. No particular set of tokens is significant; it is the data as a whole, reflected in the parameters, which enables the capabilities of a large language model. A highly sophisticated LLM trained on a large corpus

might have over a trillion parameters and be trained on trillions of tokens. *See, e.g.*, NYT Complaint ¶ 91 (alleging that GPT-4 was trained on “13 trillion tokens” and contains “1.8 trillion parameters”). It is this extensive network of semantic connections that allows the model to generate natural-language text by predicting the next likely word in a wide variety of contexts. *See id.* ¶¶ 75–77.

The resulting LLMs are sometimes referred to as “general purpose” or “foundational” AI models because their language capabilities have many applications, ranging from the mundane of drafting emails and writing code to the extraordinary of curing cancer and national defense. The user directs the LLM’s performance through user-selected “prompts” to which the model responds. Microsoft has long believed that LLMs “can so deeply absorb the nuances of language, grammar, knowledge, concepts, and context that [they] can excel at multiple tasks,” with a multitude of applications to improve people’s lives. Jennifer Langton, *Microsoft announces new supercomputer, lays out vision for future AI work*, MICROSOFT (May 19, 2020), news.microsoft.com/source/features/ai/openai-azure-supercomputer (cited at NYT Complaint ¶ 70 n.8). Recognizing the potential risks associated with this powerful technology, Microsoft has also used its leading role to urge the need for a responsible approach to deployment, routinely stressing the importance of “responsible AI and AI safety” from the highest levels of the company. *See* NYT Complaint ¶¶ 71, 93.

OpenAI has developed several versions of its foundational LLMs, built on its “Generative Pre-training Transformer” or GPT. *Id.* ¶¶ 11, 58–59. “OpenAI became a household name upon the release of ChatGPT,” a “text-generating chatbot” powered by an underlying GPT model, that “given user-generated prompts, can mimic human-like natural language responses.” *Id.* ¶ 61. Since 2019, Microsoft has invested in OpenAI to bring GPT-based products to the

public. *Id.* ¶ 66. Microsoft has provided technological infrastructure that OpenAI has used to train its LLMs, including building a supercomputing system to shoulder the immense workload of training an LLM. *Id.* ¶ 70. The Microsoft-OpenAI relationship has been instrumental in bringing the immense promise of LLMs to the public. *Id.* ¶ 72.

**B. Plaintiffs' Claims For Direct and Secondary Copyright Infringement Against Microsoft.**

Seemingly as a result of this public effort, the Plaintiffs in these cases sued Microsoft, as well as OpenAI, for copyright infringement in connection with OpenAI's GPT models. In the Consolidated Class Actions, the Individual Plaintiffs allege that Microsoft is secondarily liable for OpenAI training its LLMs on authors' fiction and non-fiction books. Class Complaint ¶¶ 2, 6, 96–130. In the Consolidated News Cases, the News Plaintiffs make the same allegations, but also accuse Microsoft of infringement in connection with products that integrate its Bing search technologies with OpenAI's foundation models. *See, e.g.*, NYT Complaint ¶¶ 2–5, 82–126. All of the cases feature variations of these copyright infringement theories in both direct and secondary (*i.e.*, contributory and vicarious) counts against Microsoft. Class Complaint ¶¶ 412–429; NYT Complaint ¶¶ 158–180. For example, The Times claims that Microsoft is secondarily liable for both OpenAI's development of its LLMs and for any allegedly infringing output resulting from end users' use of its generative AI tools. NYT Complaint ¶¶ 174–180.

Microsoft will defend these claims, *inter alia*, on the ground that the training of LLMs is a fair use of copyrighted material, and that the substantial noninfringing uses of the accused products that integrate OpenAI's LLMs with Bing search plainly vitiate any claims of liability for end-user infringement. *See* AG ECF 74 (Microsoft's Answer) at 52 (Second Defense—Fair

Use, Third Defense—Substantial Noninfringing Uses).<sup>2</sup> In determining whether a use is a “fair use” courts rely on a non-exhaustive set of factors: (1) “the purpose and character of the use,” (2) “the nature of the copyrighted work,” (3) “the amount and substantiality of the portion used in relation to the copyrighted work as a whole,” and (4) “the effect of the use upon the potential market for or value of the copyrighted work.” 17 U.S.C. § 107.

### **C. Microsoft’s Discovery Requests At Issue In The Underlying Motions.**

Amongst the discovery it served, Microsoft sought the following key evidence relevant to its defenses of fair use and substantial noninfringing uses: (1) evidence concerning the alleged economic effects of the technology at issue on the copyrighted works, and (2) evidence concerning the Plaintiffs’ efforts to develop and their uses of such technology.

***Economic Effects of Technology (Or Lack Thereof).*** Microsoft served requests on both The Times and Individual Class Plaintiffs designed to elicit evidence about the markets for and value of Plaintiffs’ works, whether any economic harm has come to those works as a result of Defendants’ products, or whether any putative decline in the economic performance of those works might be attributable to other causes. Microsoft’s requests include documents regarding: (1) The Times’s periodic financial reports, budgets, strategic and other business plans, financial projections and metrics, and periodic management reports including Board presentations regarding the financial health of The Times showing effects on revenue streams over time (NYT ECF 315-2, at 8–10 (RFP Nos. 8–9)); (2) effects on NYTimes.com domain web traffic from non-party generative AI technology (NYT ECF 315-1, at 20–21 (RFP No. 101)); (3) changes in advertising revenue from NYTimes.com webpage views and the reasons underlying same (*id.*, at 14–15 (RFP No. 89)); and (4) reasons for The Times’s subscription losses including consumer

---

<sup>2</sup> Microsoft’s Motions to Dismiss remain pending in the News cases (*see, e.g.*, NYT ECF 64), but it will interpose the same defenses in those cases once it answers.

surveys and other analyses (*id.*, at 8–10 (RFP Nos. 75–77, 80)). Microsoft sought from the Individual Class Plaintiffs any market assessments for their works, including the potential market, and valuation analyses for the copyrighted works-in-suit, both before and after the release of LLMs. AG ECF 263-1 and 263-2, at 11–14 (RFP Nos. 10–12).

***Development and Uses of LLM Technology.*** Microsoft also served discovery requests designed to elicit evidence regarding The Times’s efforts to develop, and all Plaintiffs’ uses of, generative AI tools. Based on public reporting and some documents already produced, it is apparent that The Times did at least attempt to develop its own large language model. An internal Slack chat produced by the Times indicated that The Times built “an internal ChatGPT equivalent”:

JS	Jeff Sisson	11/15/2023, 5:38 PM
<p>so wait, from the XFun all-hands just now...there's an internal ChatGPT equivalent that's been built? and a new policy that we're rolling out which means developers shouldn't use the OpenAI ChatGPT (or similar LLM) for anything, from now on?</p>		
> GM	Gaby Marraro	11/15/2023, 5:38 PM
<p>some details: &lt;<a href="https://nytimes.slack.com/archives/C050XDL1754/p1700069570871469">https://nytimes.slack.com/archives/C050XDL1754/p1700069570871469</a>&gt;</p>		

NYT ECF 320-1 (Ex. C). Additionally, The Times announced a “Gen-AI powered ad targeting solution” for advertisers, enabling its advertising customers to use generative AI in their ad campaigns. *See New York Times Advertising launches BrandMatch Out of Beta*, N.Y. TIMES (July 24, 2024), <https://www.nytco.com/press/new-york-times-advertising-launches-brandmatch-out-of-beta/>. Accordingly, Microsoft sought documents regarding The Times’s efforts to train, fine-tune, or otherwise develop its own generative AI tool(s). *See* NYT ECF 321–1, at 24–25 (Ex. A, RFP Nos. 31–32); NYT ECF 321–2, at 21, 23–24 (Ex. B, RFP Nos. 102, 106).<sup>3</sup>

Similarly, it is apparent both from public statements and from documents produced by The Times that its employees are making use of LLMs as they go about their work of creating

---

<sup>3</sup> NYT ECF 321 is the public, redacted version of NYT ECF 320.

and publishing copyrighted content. One reporter described Defendants' technology as "the best research assistant [he has] ever had." Hard Fork Ep. 77, *Can A.I. Save the Middle Class?*, YOUTUBE (April 5, 2024), <https://www.youtube.com/watch?v=Xp1Bu-ssFVg>. Rather than seek every document in every email or file about The Times's reporters' uses of generative AI, Microsoft instead served a specific and narrowly tailored request for the results of an internal survey conducted by The Times about how its employees are using the technology. NYT ECF 320-2, at 25-26 (Ex. B, RFP No. 108). Documents already produced revealed the existence of this survey:

From: **Simon Hicks** <[simon.hicks@nytimes.com](mailto:simon.hicks@nytimes.com)>  
 Date: Thu, Apr 6, 2023 at 6:12 PM  
 Subject: ACTION REQUESTED - Generative AI at NYT  
 To: Data Platform Mission - All <[data-platform.all@nytimes.com](mailto:data-platform.all@nytimes.com)>

TL;DR; Yet another fun google form to fill out! This time it's about Generative AI tools like Chat GPT. As usual it's pretty short and your input would be super, super valuable!

Hi folks,

As some of you may be aware, we've kicked off a workstream to help us empower our teams to use, explore and experiment with the various generative AI tools and capabilities that are currently available.

In order to create a set of guidelines that both encourage exploration and creativity, while protecting the company, we first want to hear from you about the tools you're interested in using and why. We've created this brief form here that we're hoping you can take 5-10 minutes to fill out. This will help us as we work toward an allowlist of use cases and technologies to help guide us all through this exciting time.

NYT ECF 320-2 (Ex. D).

Microsoft also requested that the Authors Guild and Individual Plaintiffs produce documentation regarding their own use of ChatGPT by collecting the chats from their OpenAI accounts. AG ECF 263-1, at 20 (Ex. A, RFP No. 23); AG ECF 263-2, at 21 (Ex. B, RFP No. 23). Again, there is no question that such information exists: Author Plaintiffs' counsel have conceded in meet and confers that the named Plaintiffs have made use of ChatGPT and have

even offered to provide their usernames for collection of data. AG ECF 274 at 2. This offer is of no use to Microsoft, however, as it has no access to these customer records.

As to each of these categories of requests, Plaintiffs either did not challenge the relevance, or implicitly conceded any such argument in their oppositions to Microsoft’s letter briefs. NYT ECF 332, 341; AG ECF 274. Nevertheless, Plaintiffs curtailed or outright refused production of the requested documents.

#### **D. The Orders At Issue.**

Judge Wang did not issue any substantive orders addressing the arguments Microsoft made in its letter motions. Instead, Microsoft’s requests were denied in a series of four Minute Orders (NYT ECF 351, 354, 355; AG ECF 289) that incorporated by reference an earlier Order directed to a letter brief filed by OpenAI. *See, e.g.*, NYT ECF 351 (“ORDER denying [315] Letter Motion to Compel for the reasons stated in ECF 344.”). The issues before the Court in OpenAI’s letter that led to ECF 344, however, did not include Microsoft’s requests for discovery, nor any of the Microsoft-specific reasons and legal arguments that such discovery should have been allowed.

While there was some overlap in subject matter between OpenAI’s requests and Microsoft’s insofar as Microsoft’s requests concerned The Times’s own use of generative AI, Microsoft’s requests at issue were both more targeted in seeking that information and also sought other highly relevant data regarding economic effects not covered by OpenAI’s motion. To begin with, OpenAI’s letter brief did not concern discovery related to the substitution inquiry under the fourth factor of fair use, which Microsoft sought in NYT ECF 315 & AG ECF 263. Also, Microsoft offered a number of legal arguments that were not made by OpenAI in defending the relevance of Plaintiffs’ development and use of generative AI tools, including: (1) the significance of the findings in *Google v. Oracle* regarding evidence of a copyright plaintiff’s

efforts to develop the technology at issue (NYT ECF 320/321), (2) the transformative character and purpose of technological tools under the fair use first factor (NYT ECF 320/321 & AG ECF 263) and (3) the substantial noninfringing uses defense (NYT ECF 320/321 & AG ECF 263).

Despite the difference in scope of the requests at issue and the different legal arguments, Microsoft's motions were all denied under the Order's reasoning that:

Each of [the four fair use] factors requires scrutiny of a defendant's purported use of the copyrighted work(s), and whether that defendant's use may constitute 'fair use' under the Act. The factors do not require a court to examine statements or comments a copyright holder may have made about a defendant's general industry, whether the copyright holder has used tools in the defendant's general industry, whether the copyright holder has admitted that other uses of its copyrights may or may not constitute fair use, or whether the copyright holder has entered into business relationships with other entities in the defendant's industry.

NYT ECF 344 at 2-3 (emphasis original).

## ARGUMENT

The Order at issue (ECF 344, as adopted by reference in minute orders NYT ECF 351, 354, and 355 and AG ECF 289) is unfortunately contrary to law in a number of respects, resulting largely from its failure to separately consider Microsoft's distinct discovery requests and arguments.<sup>4</sup> By refusing Microsoft discovery on basic issues such as fourth factor harm from substitution, and development and uses of the technology at issue in the ordinary course of business, ECF 344 is contrary to well established case law under the Copyright Act regarding the

---

<sup>4</sup> Pursuant to Rule 72(a), "the district judge ... must consider timely objections to the magistrate judge's decision and modify or set aside any part of the order that is clearly erroneous or is contrary to law." *Hedgeye Risk Mgmt., LLC v. Dale*, No. 21-CV-3687 ALC RWL, 2023 WL 4235768, at \*1, \*8 (S.D.N.Y. June 28, 2023) (sustaining objection to order denying motion to compel) (internal quotations and citations omitted); see also *Borjas v. N.Y.C. Dept. of Educ.*, No. 23-cv-10829, 2024 U.S. Dist. LEXIS 219579, at \*2 (S.D.N.Y. Dec. 3, 2024). A discovery order will be found "clearly erroneous if the reviewing court is 'left with the definite and firm conviction that a mistake has been committed.'" *Borjas*, 2024 U.S. Dist. LEXIS 219579, at \*2 (quoting *Easley v. Cromartie*, 532 U.S. 234, 242 (2001)). "A discovery order is 'contrary to the law' when it 'fails to apply or misapplies relevant statutes, case law or rules of procedure.'" *Hedgeye*, 2023 WL 4235768, at \*8 (quoting *Weiss v. LaSuisse*, 161 F. Supp. 2d 305, 321 (S.D.N.Y. 2001)).

scope of Microsoft’s defenses.<sup>5</sup> First, it is contrary to law insofar as it denied Microsoft access to evidence exclusively within Plaintiffs’ hands regarding the economic effects on them (or lack thereof) of the technology at issue relevant to the issue of substitution under the fourth factor fair use analysis—discovery that was not the subject of OpenAI’s letter brief at all, nor even squarely addressed in ECF 344. Part I, *infra*. Second, ECF 344 is also contrary to law in restricting the production of evidence—again, entirely within Plaintiffs’ hands—relating to their development and use of generative AI tools; on these requests, ECF 344 both fails to address several controlling points of law put forward by Microsoft and adopts an insupportably narrow conception of the evidence relevant to the fair use defense. Part II, *infra*.

Microsoft appreciates the hard work that Magistrate Judge Wang is devoting to these cases and does not take lightly the decision to file an objection. This Objection is not a minor quibble. ECF 344, when applied to refuse Microsoft’s discovery requests, goes to the heart of the evidence that can be adduced regarding key defenses in these cases. The big picture result of applying the reasoning of ECF 344 broadly across a number of issues—and the crux of the necessity of this Objection—is tantamount to preventing Microsoft from fully presenting the merits of its most important defenses in these cases of first impression involving revolutionary technology.

## **I. THE ORDER ERRONEOUSLY REFUSES DISCOVERY INTO ECONOMIC EFFECTS RELEVANT TO A KEY FOURTH FACTOR ISSUE: SUBSTITUTION.**

The *sine qua non* of the fourth factor of fair use is economic harm through *substitution*.

*Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 578-79 (1994) (“The central purpose of this

---

<sup>5</sup> Microsoft also notes that ECF 344 was decided on the basis of a single exchange of 3-page letters, and then applied without comment to Microsoft’s motions which were also limited to 3-page letters, without opportunity for full motion briefing on any of the issues. Microsoft also was not afforded an opportunity to argue its motions.

investigation is to see, in Justice Story's words, whether the new work merely 'supersede[s] the objects' of the original creation, ('supplanting the original')") (internal citations omitted); *Harper & Row, Publ'rs. v. Nation Enters.*, 471 U.S. 539, 568 (1985) (The substitution inquiry concerns whether the "use [] supplants any part of the normal market for a copyrighted work."); *Authors Guild v. Google, Inc.*, 804 F.3d 202, 223 (2d. Cir. 2015) (The fourth factor "focuses on whether the copy brings to the marketplace a competing substitute for the original, or its derivative, so as to deprive the rights holder of significant revenues because of the likelihood that potential purchasers may opt to acquire the copy in preference to the original.") (emphasis original); *see also Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 100 (2d Cir. 2014).

Thus, where plaintiffs claim to be suffering harm to the value of their works, the fourth factor inquiry requires courts to ascertain whether that harm is coming from substitution or from some cause *other than* substitution. "[A] potential loss of revenue is not the whole story. We here must consider not just the amount but also the source of the loss." *Google LLC v. Oracle Am., Inc.*, 593 U.S. 1, 35 (2021). Even the suppression of demand as a result of the defendant's copying is not remediable under the Copyright Act where that suppression is caused by something other than displacement in the market through substitution. *Campbell*, 510 U.S. at 591-92. Only substitution counts as relevant harm for assessing the fourth factor of fair use. *NXIVM Corp. v. Ross Inst.*, 364 F.3d 471, 481–82 (2d Cir. 2004) (focus is "on whether defendants are offering a market substitute for the original" and "whether the secondary use usurps the market of the original work"). Cognizable market harm is limited to market substitution. *Authors Guild v. HathiTrust*, 755 F.3d at 96, 99-100.

Relevant evidence for this inquiry encompasses all "traditional," "reasonable," or "potential markets." *Am. Geophysical Union v. Texaco Inc.*, 60 F.3d 913, 930 (2d Cir. 1994)

(“*Texaco*”); *see also Bill Graham Archives v. Dorling Kindersley Ltd.*, 448 F.3d 605, 614 (2d Cir. 2006). To help define those markets and assess the causal question of substitution, Microsoft sought to investigate whether and why The Times had suffered any (1) subscription losses, including data from user surveys about *all* reasons why they cancelled their subscription; (2) changes to advertising revenue over time and any analyses of the causes of such changes; or (3) changes in web traffic to The Times’s websites over time and any analyses of the causes of such changes. Microsoft also sought standard economic and financial discovery in the form of periodic financial reports, budgets, strategic and other business plans, financial projections and metrics, and periodic management reports including Board presentations regarding the financial health of The Times showing effects on revenue streams over time. *See* NYT ECF 315.<sup>6</sup> What The Times’s officers report to its Board and predict about its financial performance will reflect the most pertinent understanding of factors affecting the value of its works. If The Times intends to argue that generative AI tools like ChatGPT or Copilot are substituting for and replacing the market for their works under the fourth factor, Microsoft should be able to take discovery of The Times own explanations for any decline (or increase) in the popularity and value of their works. *Only* The Times has this understanding of its business, and the Court’s orders stop Microsoft from acquiring this information.

Crucially, these categories and subject matter of discovery were not before the Court in OpenAI’s original letter brief at all, nor did OpenAI advance any argument that the evidence it

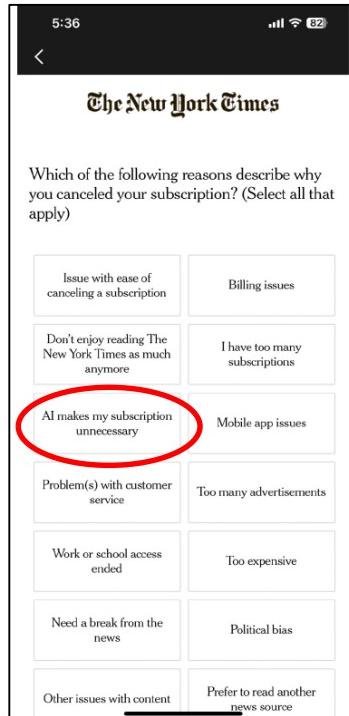
---

<sup>6</sup> Microsoft also sought information from the Authors Guild regarding any valuation or market analyses it has performed of the Class Works and for the Individual Plaintiffs to request certain financial information from their literary agents and publishers. AG ECF 263. The Individual Plaintiffs have agreed to request certain relevant market and valuation documentation from their literary agents and publishers such that this issue is resolved, and Court intervention is no longer necessary. To the extent the Plaintiffs fail to adhere to this agreement, Microsoft reserves the right to seek relief from the Court. Additionally, in response to the Authors Guild’s objection that the requests served on it were more narrowly focused and did not extend to the entirety of the Class Works, Microsoft has served additional requests on the Authors Guild.

sought was relevant to the issue of fourth factor substitution. *See* NYT ECF 236. For that reason alone, it was error for the Court to deny Microsoft’s letter motion (NYT ECF 315) “for the reasons stated in ECF 344,” as the discovery at issue, as well as Microsoft’s arguments defending it, differed considerably from those considered in the prior Order.

To the extent the reasoning of the Order can be read broadly to somehow encompass the subject matter of fourth factor harm from economic substitution, the necessary causal inquiry cannot possibly be accomplished by focusing exclusively on *Defendants’* actions, as ECF 344 holds. If the focus is only on the Defendants, the Court would learn little or nothing about the markets for Plaintiffs’ works, the performance of Plaintiffs’ works in those markets, and other economic factors affecting the performance of Plaintiffs’ works in those markets after the Defendants’ products were introduced, such as the well-known increase in reading during the pandemic. *See* Bureau of Labor Statistics, *U.S. consumers spent more on recreational reading during COVID-19 pandemic than before*, U.S. DEP’T OF LABOR, (Oct. 27, 2022), <https://www.bls.gov/opub/ted/2022/u-s-consumers-spent-more-on-recreational-reading-during-covid-19-pandemic-than-before.htm> (visited December 17, 2024).

Moreover, subscription, advertising, and referral revenue of the type earned by the News Plaintiffs through monetizing their works depend upon the actions and “eyeballs” of readers. Such revenue is inherently tied up with the actions of others, and any harmful effects from generative AI on such revenue streams could only be assessed by looking at how readers are behaving in light of the advent of generative AI. The best source of evidence about how the News Plaintiffs’ customers are behaving is from the News Plaintiffs. Indeed, The Times conducted a survey of users where it asked them to specify their reasons for cancellation, and one of the reasons it proposed was that “AI makes my subscription unnecessary”:



NYT ECF 341-2 (Ex. 2). While even The Times concedes it must produce this survey, it refuses to produce all other surveys regarding its consumer behavior unless they *specifically mention* generative AI. Yet, The Times's inquiries of the reasons for its customers' behavior, relied upon by it in the ordinary course of its business, are plainly the most relevant evidence to assessing the causes of any alleged losses.

Accordingly, Microsoft's requested documents are directly relevant to the fourth factor harm inquiry. Indeed, Microsoft is obligated in the first instance to come forward with such evidence. *Campbell*, 510 U.S. at 590 (reversing summary judgment because defendant failed to meet burden of production on fourth factor). It is difficult enough to prove the negative of a lack of harm. It would be fundamentally unfair, as well as at odds with the adversarial process and the scope of discovery permissible under Rule 26, were Microsoft to be limited in doing so only with evidence that Plaintiffs themselves decide to use to support their case. *United States CFTC v. Parnon Energy, Inc.*, 593 Fed. Appx. 32, 36 (2d Cir. 2014) ("Relevance to the subject matter

under Rule 26 is ‘construed broadly to encompass any matter that bears on, or that reasonably could lead to other matter that could bear on, any issue that is or may be in the case.’’’). And yet, this is precisely the effect of ECF 344.

Microsoft is entitled to seek evidence of alternate explanations. *See, e.g., Granite State Ins. Co. v. Clearwater Ins. Co.*, No. 09-cv-10607 (RKE), 2012 WL 1520851, at \*3–4 (S.D.N.Y. Apr. 30, 2012) (plaintiff’s contention that documents relevant to bad faith defense were irrelevant based on its contested construction of the scope of this defense was not a sufficient basis to deny discovery). Discovery is not just a “one-way street designed to allow plaintiffs to collect evidence in support of their claims. It is also a mechanism for defendants to accumulate evidence to defend themselves and to test the evidence of their opponents.” *Fioranelli v. CBS Broad., Inc.*, No. 15-cv-952 (VSB), 2019 WL 1059993, at \*5 (S.D.N.Y. Mar. 6, 2019); *see also Granite State*, 2012 WL 1520851, at \*3-4 (overruling objection and upholding magistrate judge’s order compelling discovery relevant to defendant’s defenses).

To the extent ECF 344 discusses concepts of economic harm at all, it seems to find that only a licensing market for the specific use at issue is relevant in these cases. NYT ECF 344 at 4-5. Such evidence might or might not be pertinent. *See HathiTrust*, 755 F.3d at 100 (“it is irrelevant that the [defendants] might be willing to purchase licenses in order to engage in this transformative use”). But even if such evidence were relevant, it is far from the *only* evidence to consider in connection with the fair use fourth factor. Microsoft is entitled to establish a lack of harm in the primary and other established markets for the works in order to establish a lack of substitution and to meet its burden under *Campbell*, and in order to do so it is entitled to the requested discovery on the economic performance of the News and Individual Plaintiffs’ works in those markets. If the evidence shows, as Microsoft believes it will, that the accused

technology had no effect on the existing markets for Plaintiffs' works, then such evidence may well be dispositive in demonstrating a lack of substantial harm to the value of the works.

It was contrary to well established law on fair use to limit discovery on economic effects to only a specific licensing market for model training. Because the requested economic evidence regarding substitution is of a type exclusively in the hands of Plaintiffs, this discovery ruling also has the effect of improperly limiting the merits of the fair use defense, thus also leaving the definite and firm conviction that a mistake has been made. Accordingly, the Orders denying the relief requested in NYT ECF 315 (Order, NYT ECF 351) and AG ECF 263 (Order, AG ECF 289) should be set aside under Rule 72(a).

## **II. THE ORDER IMPROPERLY RESTRICTS THE PRODUCTION OF EVIDENCE RELATED TO THE DEVELOPMENT AND USE OF GENERATIVE AI BY PLAINTIFFS.**

With respect to evidence relating to the Plaintiffs' development and use of Generative AI tools, the Order as applied to Microsoft's letter motions is contrary to law in three ways. *First*, the Order did not address Microsoft's argument that discovery into The Times's attempts to develop its own generative AI tools is relevant to whether it suffered fourth factor harm under the Supreme Court's recent decision in *Google v. Oracle*. *Second*, in relying upon an inapt video game analogy, ECF 344 fails to consider the purpose of the use in creating a general-purpose language *tool*; as a result, it improperly restricts Microsoft from eliciting key evidence from the Plaintiffs as to how they are using these tools. *Third*, the Order fails to address Microsoft's separate legal argument establishing the relevance of use evidence to its substantial noninfringing uses defense.

### **A. The Order Fails To Address The Application Of *Google v. Oracle* To The Relevance Of Evidence Of The Times's Training Of Its Own Model.**

In denying NYT ECF 320/321 "for the reasons stated in ECF 344," the Order erred in

failing to address the application of the Supreme Court’s decision in *Google v. Oracle* to the evidence sought by Microsoft regarding The Times’s efforts to train its own LLM. As discovery has already revealed, The Times undertook an effort to develop LLM technology:

JS	Jeff Sisson	11/15/2023, 5:38 PM
so wait, from the XFun all-hands just now...there's an internal ChatGPT equivalent that's been built? and a new policy that we're rolling out which means developers shouldn't use the OpenAI ChatGPT (or similar LLM) for anything, from now on?		
> GM	Gaby Marraro	11/15/2023, 5:38 PM
some details: < <a href="https://nytimes.slack.com/archives/C050XDL1754/p1700069570871469">https://nytimes.slack.com/archives/C050XDL1754/p1700069570871469</a> >		

NYT ECF 320-1 (Ex. C). Accordingly, Microsoft requested documents concerning those efforts. *See generally* NYT ECF 320; *see also* NYT ECF 321-1, at 24-25 (Ex. A, RFP Nos. 31-32); NYT ECF 321-2, at 21, 23-24 (Ex. B, RFP Nos. 102, 106). In seeking to compel this evidence, Microsoft relied upon the Supreme Court’s decision in *Google v. Oracle*, where the Supreme Court discussed at length the evidence regarding the importance to the fourth factor of the plaintiff in that case’s efforts to develop exactly the technology at issue. *See Google*, 593 U.S. at 35-36 (plaintiff’s inability to develop the technology was directly relevant to whether it suffered fourth-factor harm); NYT ECF 320 at 1-2.

Public (and internally produced) documentation shows that The Times has certainly tried to train and develop its *own* generative AI tools. Whether and how it has been able to do so is directly relevant to the question of whether Microsoft’s tools could somehow “substitute” for that of The Times. If The Times was unable to train a tool using only its own content as tokens during the training process, such an outcome would strongly suggest a lack of value of its works to the training of LLMs and correspondingly a lack of harm. If The Times used the copyrighted material of others as part of a training dataset to generate the necessary quantity of tokens in its efforts to train an LLM, such activity would strongly suggest that the industry either views such copying for the purposes of AI training as a fair use, or at a minimum that such “unauthorized”

uses are not willful infringement. If The Times was unable to train a model despite its unfettered access to the more than 10 million works it asserts were infringed in this case, then its failure strongly suggests that it is not particular tokens that lend value to generative AI models but the technological prowess of the Defendants in developing and implementing such models.

All of these scenarios and more are highly relevant to consideration of fair use under the *Google v. Oracle* decision, where the Supreme Court highlighted that plaintiff had made “many efforts to move into the mobile phone market” which were ultimately “unsuccessful.” *Google*, 593 U.S. at 36. That the plaintiff had tried and failed to enter the market at issue was directly relevant to the Supreme Court’s fourth factor fair use inquiry, and in particular the question of “the effect of the use upon the potential market for or value of the copyrighted work.” *Id.* at 35-36. The inability of the plaintiff in that case to exploit the market resulted in the fourth factor weighing against it. Yet here, the Order refuses the very discovery that would be necessary to make such a showing. In rejecting the requested discovery, ECF 344 does not even mention the controlling passage of this Supreme Court opinion. And, in fact, the core holding of ECF 344 that fair use is not concerned with Plaintiffs’ actions is directly contrary to this very analysis undertaken by the Supreme Court.

Microsoft explicitly relied upon this specific passage in the *Google* decision; OpenAI did not mention it. NYT ECF 320 at 2 (“Any failed efforts by The Times to develop its own Generative AI system using its works would undermine its claims of harm. *Google*, 593 U.S. at 36 (failure to develop technology weighed against harm in analysis of fourth factor.”). The Order only addresses those parts of the *Google* decision raised by OpenAI; the failure to address Microsoft’s separate legal argument results in a decision contrary to controlling Supreme Court authority. The Court should set aside NYT ECF 354/355 and compel production of documents

regarding The Times's efforts to train its own large language models for the additional and independent reason that *Google* compels its relevance to the fourth factor.

**B. The Order Improperly Narrows The Scope Of Relevant Uses For The First Factor.**

The Order also improperly refused discovery into evidence regarding the Plaintiffs' use of generative AI tools, holding that only the Defendants' uses of the works in creating the tools were relevant to the question of transformative use and that the Plaintiffs' subsequent uses of the resulting tools have no bearing whatsoever on the question of fair use. NYT ECF 344 at 2-3. This undue restriction on the scope of evidence relevant to the first factor of fair use is contrary to law.

The purpose of Defendants' copying is to create technological tools that support and enable a wide variety of other activities. Whether the purpose of creating those technological tools is a transformative use of the underlying works is a fundamental part of the first factor fair use inquiry. Thus, it is not just the immediate steps of transformation in creating the large language models that are at issue in assessing transformative use, but also how people interact with the result of the copying—this is particularly so when the purpose of the use at issue is to create a technological tool for use by others. *See Google*, 593 U.S. at 40 (analyzing use of program developers “to put their accrued talents to work in a new and transformative program”); *Authors Guild v. Google*, 804 F.3d at 214-220, 225 (analyzing use “by the public” of search and snippet view functions); *Authors Guild v. HathiTrust*, 755 F.3d at 97-98, 101-102 (analyzing potential use by scholars of full-text book searching); *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146, 1155-57, 1168 (9th Cir. 2007) (analyzing search by users employing thumbnail images); *Fox News Network, LLC v. TVEyes, Inc.*, 43 F. Supp. 3d 379, 392-393, 397 (S.D.N.Y.

2014) (analyzing subscriber uses of indexing and clipping services).<sup>7</sup>

This is why the video game analogy employed in ECF 344 is inapt. Simple re-use of a visual or textual work in another entertainment medium is a vastly different proposition than transforming it into a functional tool with capabilities far beyond the original. *Compare Google*, 593 U.S. at 30-31 (finding Google’s use of copyrighted material, which was “to create new products” which “expand[ed] the use and usefulness” of smartphones, was “consistent with that creative ‘progress’ that is the basic constitutional objective of copyright itself, and thus transformative) with *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 547 (2023) (“*Warhol*”) (finding AWF’s commercial use of the copyrighted work to illustrate a magazine about Prince too similar to the work’s typical use, and thus not transformative despite aesthetic changes that had been made).

The Order shuts down Microsoft’s inquiry into highly salient uses reflected in a survey of employees conducted by The Times, casting the discovery as “a referendum on the benefits of GenAI, on Plaintiff’s business practices, or about whether any of Plaintiff’s employees use Gen AI at work.” NYT ECF 344 at 5. But Microsoft’s discovery was focused and specific: the results of an internal survey conducted by The Times of how its employees use generative AI tools. Publicly available material suggests that this survey is likely to produce a goldmine of transformative uses by the Plaintiff itself. As just one example, a Times reporter declared—on one of The Times’s official podcasts—that generative AI was the best research tool he has ever used. Hard Fork Ep. 77, *Can A.I. Save the Middle Class?*, YOUTUBE (Apr. 5, 2024) <https://www.youtube.com/watch?v=Xp1Bu-ssFVg>. It is therefore apparent that production of a Times’s survey and its results is one of the most efficient and proportionate means of obtaining

---

<sup>7</sup> This evidence is also relevant to Microsoft’s separate substantial noninfringing uses defense. See Part I.D, *infra*.

ordinary course of business evidence of The Times's uses of the technology at issue.

The Order is also erroneous in suggesting that the broader uses of LLM technology (which range from curing cancer to national defense) are not cognizable in the fair use analysis. The Order seemed to contemplate that only those purposes relevant to *copying* should be considered. NYT ECF 344 at 5 (citing *Warhol*, 598 U.S. at 528).<sup>8</sup> While there is a debate about the relevant scope of uses, *see generally* Matthew Sag, *Fairness and Fair Use in Generative AI*, 92 Fordham L. Rev. 1887 (2024), Microsoft's requested survey evidence about how The Times uses the technology is specifically relevant to Copyright Act purposes, and thus relevant to the first factor fair use inquiry. Wherever the Court ultimately draws the line in connection with summary judgment and/or trial, what Microsoft seeks is relevant because it is a survey about how the technology is used by *employees in an organization whose purpose is to engage in the creation and publication of copyrighted material*. Those employees have made public statements that they are using the technology for pro-copyright purposes. Accordingly, Microsoft's discovery directly addresses transformation for pro-copyright uses that are relevant under even the most miserly interpretation of first factor purposes, asking whether and to what extent The Times is using generative AI tools “for purposes such as criticism, comment, news reporting, … scholarship, or research,” which “is not an infringement of copyright.” 17 U.S.C. § 107; *see Texaco*, 60 F.3d at 923 (“making some contribution of new intellectual value and thereby fostering the advancement of the arts and sciences[.]”).

Indeed, the Supreme Court in *Google v. Oracle* described how the first factor inquiry involves “asking whether the copier’s use fulfill[s] the objective of copyright law to stimulate

---

<sup>8</sup> The Order also appears to focus its discussion of benefits on “public benefits” in the context of the fourth, not first, fair use factor. NYT ECF 344 at 3-5 (discussing *Google* and *Warhol*). While still misguided, it is also clear that the Court was not focused on the arguments Microsoft advanced on this point relating to the relevance to the first factor of fair use.

creativity for public illumination.” *Google*, 593 U.S. at 29 (internal citation omitted). In these technological tool cases, the consideration of stimulating creativity inherently involves looking at the actions of others in using the challenged tools. Thus, in that case defendant created a new platform that could be readily used by other programmers to develop new applications for smartphones, a use that “was consistent with that creative ‘progress’ that is the basic constitutional objective of copyright itself.” *Id.* at 30-31 (*citing Cf. Feist Publications, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 349-350 (1991) (“The primary objective of copyright is not to reward the labor of authors, but ‘[t]o promote the Progress of Science and useful Arts.’”)) (quoting U.S. Const., Art. I, §8, cl. 8”)). A purpose to encourage pro-copyright uses—which the Order implicitly concedes are relevant—requires looking at the uses of others to assess the first factor fair use analysis.

Separately, the Order also appears to have excluded this evidence based on an assumption that The Times could be using other generative AI tools and that such tools “presumably were developed without copying.” NYT ECF 344 at 4. Without obtaining this discovery, we will never know which generative AI tools The Times is using or how they were trained. And the assumption that all those tools were trained without the benefit of any copyrighted material is unlikely to be correct given that the known LLMs today are all trained generally in the same way as the LLMs at issue in this case, using large volumes of material obtained from the internet. *See, e.g., Concord Music Grp., Inc. et al. v. Anthropic PBC*, No. 5:24-cv-03811 (N.D. Cal.); *UMG Recordings, Inc. et al. v. Suno, Inc. et al.*, No. 1:24-cv-11611 (D. Mass.). To the extent ECF 344 relies upon such speculation in refusing the requested discovery, it is clearly erroneous.

For all of these reasons, the Court’s Orders (NYT ECF 354/355; AG ECF 289) denying Microsoft’s letter motions to compel discovery regarding Plaintiffs’ uses of generative AI

technology are contrary to law and should be reversed.

**C. The Order Fails To Address Microsoft’s Substantial Noninfringing Uses Defense (NYT ECF 354/355 & AG ECF 289).**

Finally, discovery into Plaintiffs’ uses of LLM tools is directly relevant to the defense of substantial noninfringing uses. Plaintiffs seek to hold Microsoft secondarily liable under a theory of contributory infringement for both the actions of OpenAI in training its LLMs and for the actions of others with respect to the *outputs* of such models during use. Whether Microsoft can be held secondarily liable for the operation of OpenAI’s GPT models under a theory of contributory infringement depends in part on whether those models are capable of substantial noninfringing uses. *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 442 (1984) (use “does not constitute contributory infringement if the product is widely used for legitimate, unobjectionable purposes.”); *see also EMI Christian Music Grp., Inc. v. MP3tunes, LLC*, 844 F.3d 79, 100-01 (2d. Cir. 2016); *Cartoon Network LP, LLLP v. CSC Holdings, Inc.*, 536 F.3d 121, 133 (2d Cir. 2008).

Crucially, neither OpenAI’s briefing nor ECF 344 itself make any mention of the substantial noninfringing uses doctrine nor of any of the relevant case law. Yet the Court denied the relevant discovery sought by Microsoft without further explanation. Nothing in the “reasons stated in ECF 344” demonstrates that the Court ever considered the relevance of any discovery requests, much less Microsoft’s specific requests, or the case law relevant to this critical defense. Who better to reflect the ordinary noninfringing uses of these tools than the Plaintiffs themselves?

Additionally, it is not clear whether the models in ordinary operation are even capable of producing *infringing* outputs. The News Plaintiffs disavowed their efforts undertaken prior to litigation to demonstrate that the models commit output infringement, and instead have

demanded hundreds of thousands of dollars in account value to test the models through inspection to see if they can demonstrate what they call “regurgitation.” Surely if Plaintiffs’ uses during the artificial conditions of a litigation inspection by experts designed to produce adverse evidence are relevant, then their own *ordinary course of business* uses are even more so. How better to cross examine Plaintiffs’ experts regarding their artificially generated infringing uses than with Plaintiffs’ own noninfringing ordinary course of business uses?

This is certainly the type of evidence the Supreme Court considered in *Sony* when it evaluated declarations and other testimony from content producers that they believed that uses by the public of the Betamax were beneficial and not harmful. *Sony*, 464 U.S. at 443-47. The results of The Times’s survey regarding its employees’ uses of AI tools is similarly likely to show a lack of harm, not only because they are “the best research assistant...ever” but also because they show a wide variety of other noninfringing uses. The Individual Plaintiffs’ ChatGPT accounts are also likely to show them generating ideas, research, and other quintessentially appropriate pro-copyright uses. The fact that such evidence will dramatically skewer Plaintiffs’ cases makes it highly relevant and is not a basis for Plaintiffs to avoid such discovery.

The evidence of Plaintiffs’ uses is directly relevant to the substantial noninfringing uses defense, and the Order’s failure to address Microsoft’s defense to claims of contributory infringement was contrary to law.

### **CONCLUSION**

For the reasons presented herein, Microsoft respectfully requests that pursuant to Rule 72(a) the Court sustain its Objection and set aside NYT Orders ECF 351, 354, 355 and AG Order ECF 289, and NYT ECF 344 insofar as necessitated by the arguments set forth herein, and mandate that Plaintiffs produce Microsoft’s requested discovery in NYT ECF 315, 320/321 and AG ECF 263.

Dated: December 23, 2024

/s/ Annette L. Hurst

---

Annette L. Hurst (admitted *pro hac vice*)  
ORRICK, HERRINGTON &  
SUTCLIFFE LLP  
The Orrick Building  
405 Howard Street  
San Francisco, CA 94105  
Telephone: (415)773-5700  
ahurst@orrick.com

Christopher Cariello  
Marc Shapiro  
ORRICK, HERRINGTON &  
SUTCLIFFE LLP  
51 West 52<sup>nd</sup> Street  
New York: NY 10019  
Telephone: (212) 506-3778  
Facsimile: (212) 506-5151  
ccariello@orrick.com  
mrshapiro@orrick.com

Sheryl Koval Garko (admitted *pro hac vice*)  
Laura Brooks Najemy (admitted *pro hac vice*)  
ORRICK, HERRINGTON &  
SUTCLIFFE LLP  
222 Berkeley Street, Suite 2000  
Boston, MA 02116  
Telephone: (617) 880-1800  
Facsimile: (617) 8801-1801  
sgarko@orrick.com  
lnajemy@orrick.com

Jeffrey S. Jacobson  
FAEGRE DRINKER BIDDLE  
& REATH LLP  
1177 Avenue of the Americas  
New York, New York 10036  
Telephone: (212) 248-3191  
jeffrey.jacobson@faegredrinker.com

Jared B. Briant (admitted *pro hac vice*)  
Kirstin L. Stoll-DeBell (admitted *pro hac vice*)  
FAEGRE DRINKER BIDDLE  
& REATH LLP  
1144 Fifteenth Street, Suite 3400

Denver, CO 80202  
Telephone: (303) 607-3500  
[jared.briant@faegredrinker.com](mailto:jared.briant@faegredrinker.com)  
[kirstin.stolldebell@faegredrinker.com](mailto:kirstin.stolldebell@faegredrinker.com)

Carrie A. Beyer (admitted *pro hac vice*)  
FAEGRE DRINKER BIDDLE & REATH LLP  
320 South Canal Street, Suite 3300  
Chicago, IL 60606-5707  
Telephone: (312) 569-1000  
Facsimile: (312) 569-3000  
[carrie.beyer@faegredrinker.com](mailto:carrie.beyer@faegredrinker.com)

Elizabeth M.C. Scheibel (admitted *pro hac vice*)  
FAEGRE DRINKER BIDDLE & REATH LLP  
2200 Wells Fargo Center, 90 S. 7<sup>th</sup> Street  
Minneapolis, MN 55402  
Telephone: (612) 766-7000  
Facsimile: (612) 766-1600  
[elizabeth.scheibel@faegredrinker.com](mailto:elizabeth.scheibel@faegredrinker.com)

*Attorneys for Defendant Microsoft Corporation*